

XMLとデータベースを活用したドキュメント制作システムの概要

Outline of Database/XML-based Document Production System

吉田 樹治, 間島 誠也, 大橋 則行

Shigeji Yoshida, Seiya Majima, Noriyuki Ohashi

要旨 ドキュメントの制作・管理作業にXML, データベース, 自動組版の技術を適用したシステムを開発した。このシステムでは, ドキュメントの構造と文章を分離してデータベースで管理することで, 複数のドキュメントで文章を共有することを可能にしている。またデータベースには言語の管理機能を持たせており, 英語で作成したドキュメントは, 西語や仏語など他の言語で書き出すことができる。最終的にはデータベースから文章を抽出し, XML形式で記述されたドキュメントを生成して, 自動組版している。

Summary This system has been developed to apply XML, database and automated layout technologies to document production and management. Using this system, the structure and text of a document are contained separately in a database, enabling content to be shared among several documents. The database can also be used to manage multiple languages, generating Spanish, French and other languages from an original English document. Finally, text can be extracted from the database, XML-tagged documents created, and page layout automated.

キーワード : ドキュメント制作, 取扱説明書, 多言語展開, 翻訳管理, XML, データベース, 自動組版

1. まえがき

製品の複雑化・高機能化が進む中, 取扱説明書の制作フローにおいて, 開発日程や品質をキープしながら数多くの言語に同時展開していく難しさには想像を上回るものがある。多くの制作会社で中心的で, 広く普及しているDTP¹ (DeskTop Publishing)と称される制作方法では, 次のような点が課題としてあげられる。

- ・個々の取扱説明書ごとにデータを作成するため, 同じ機能の説明でも文章の共有

は困難である(修正がある場合は, それぞれの取扱説明書で修正する必要がある)。

- ・見た目(レイアウト)を整える組版作業に人手を介するため, 一定の編集工数が必要である。
- ・多言語に展開する場合は翻訳期間が必須であるが, 翻訳依頼文章の抽出や取り込み作業にも手間がかかる。日程に重なりがあるモデルでは二重翻訳になる傾向にあり, 翻訳資産を生かしきれない。

このような状況下での仕様変更や設計変更は、修正やチェック、文書の管理に、×言語数、×展開モデル数の工数が必要となり、効率が良いとは言い難い。

そこで、文章の共有・再利用を促進するためにデータベースを使用し、また組版作業を自動化するためにXMLを利用したドキュメント制作システムを開発した。

XML²(eXtensible Markup Language)は、拡張性に優れた情報記述形式として注目されており、各業界団体でさまざまな取り組みが行われている。しかしドキュメント制作分野において、構造化データという特徴を活かした既存アプリケーション利用例は、まだあまり見られない。このため、XMLドキュメントを制作するアプリケーションの開発から着手した。

2. システムの構成

システムのねらいは、

- ・ 構造や文章の共有，再利用
- ・ 翻訳コントロールの効率化
- ・ 組版(レイアウト)作業の自動化
- ・ 品質の均一化

である。

システム全体の構成を図1に示す。このシステムはXMLドキュメント制作アプリケーションと自動組版アプリケーションに分かれている。

XMLドキュメント制作アプリケーションの役割は大きく、

- ・ DTD³(Document Type Definition: 文書型定義)に沿って構造を組み立てる構造管理部
- ・ 文章(コンテンツ)の状態や履歴をデータベース化する文章管理部
- ・ 未翻訳の文章を翻訳依頼するために翻訳原稿の出力/入力を行う翻訳管理部
- ・ 構造と文章を結びつけてドキュメントをXML形式で出力するXML文書生成部

に分類される。

自動組版アプリケーションは生成されたXML文書を組版し、冊子という体裁に整える役割を持つ。

3. XMLドキュメント制作アプリケーション

図2にXMLドキュメント制作アプリケーションの概念を示す。ドキュメントの構造とドキュ

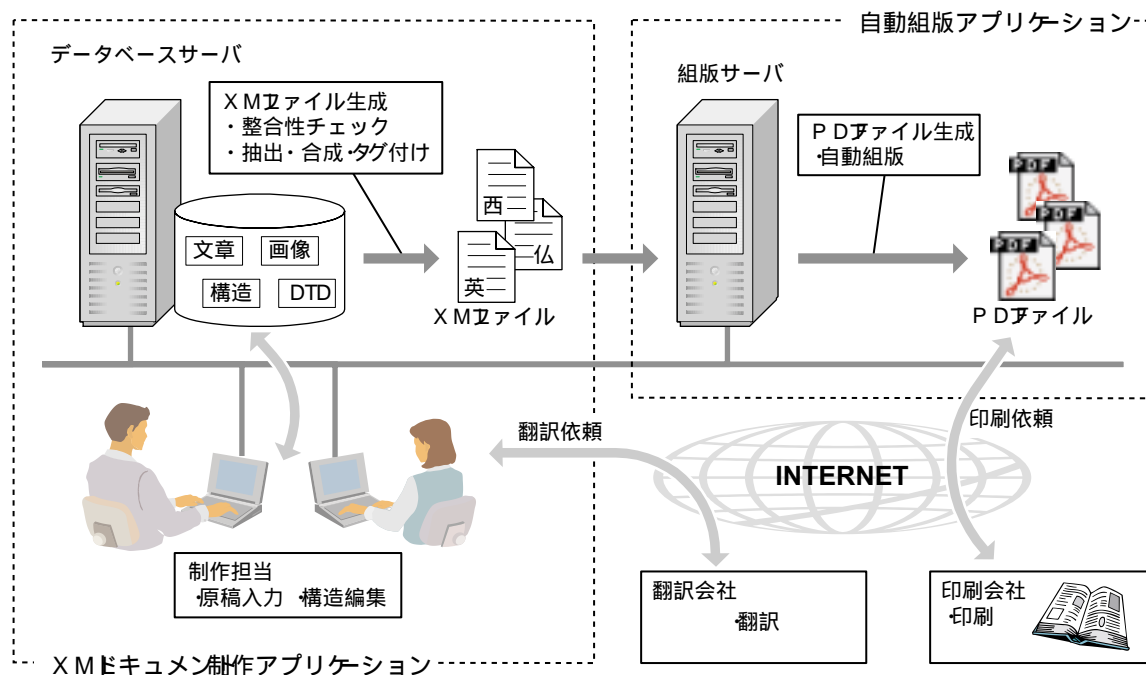


図1 システムの構成

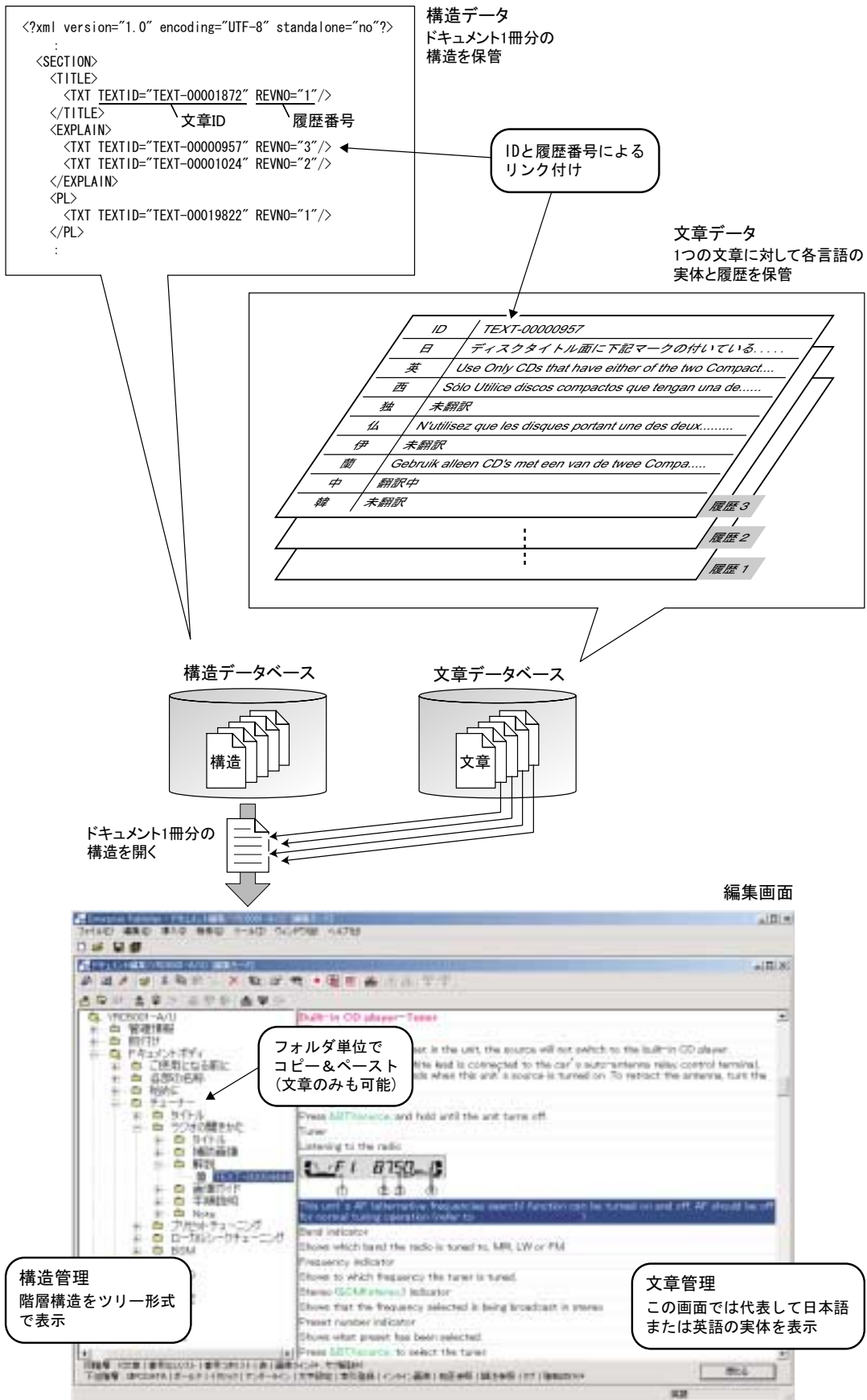


図 2 XML ドキュメント制作アプリケーションの概念

メント内で使われている1つ1つの文章を分離して、データベースで管理する。構造データには実際に入る文章(XMLの規格でいうCDATA)の代わりに、文章のIDと履歴番号を持たせることで文章データとリンク付けを行っている。

構造データをアプリケーションで開くと、左側にドキュメントの構造を表したツリー図、右側に文章実体を配置した編集画面が表示される。制作担当者はこの編集画面を通して構造や文章の追加・修正・削除を行い、編集結果をデータベースに保存する。

ドキュメント内やドキュメント間では任意のフォルダ(要素)ごとにコピー＆ペーストを行うが、構造データをコピーするだけで文章は1つのIDを参照するため、複数ドキュメントで文章を共有している状態となる。

3.1 構造管理

文章の共有率を高めるためには、ドキュメントの構造が標準化されていなければならない。そこで取扱説明書の構造をDTDとして明確に定義し、ドキュメントに新しく要素を追加する際は、DTDで定義された要素以外は選択できないようにするアシスト機能を盛り込んだ。制作担当者が他のドキュメントから要素をコピー＆ペーストした際も、DTDで定義されていない場合は色を変えて表示する。この機能を図3に示

す。この機能により誤った要素の挿入を防ぐことができ、制作担当者が構造に精通していない場合でも均一化されたドキュメントの制作が可能になる。

アシスト機能は、DTDの定義内容に情報を付加して独自形式でデータベースに登録することで実現している。さらにこの情報を利用してXPath⁴(XML Path Language)の一部の指定をGUI操作で実現しており、図4に示すように、全ドキュメントを対象に、任意の階層を辿った構造による検索が可能である。XPathはXMLの関連規格で、構造の一部を特定する際に使用される。例えばXPathでは、セクションのタイトルが「CDの聞きかた」で下位階層の「解説」という要素を抽出するのに、

```
select="SECTION[TITLE='CDの聞きかた']//解説"
```

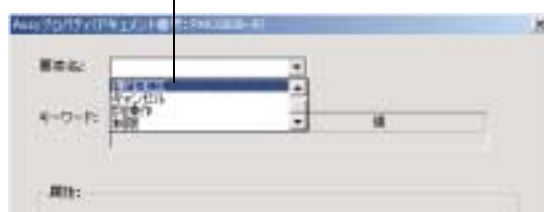
という記述をするが、これを図4に示すようなGUI操作で行うことができる。

3.2 文章管理

文章管理の概念は前述の図2に示した通りである。文章を部品としてとらえ、IDを付与してデータベースに登録する。そのIDに対して各言語の実体と修正履歴、翻訳状況などを管理する。

図5の文章編集画面には、図2で右側に表示された文章実体をダブルクリックして遷移する。新規文章の追加の場合も、IDを新たに取得

新規要素追加の際、挿入できる要素のみが表示される



コピー＆ペーストにより下位階層に誤った要素が挿入された場合、赤く表示される

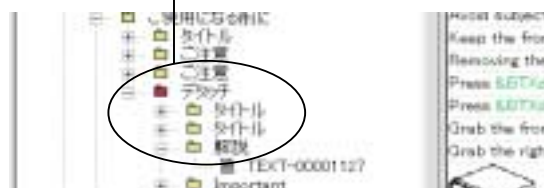


図3 DTDのアシスト機能

データベース内の全ドキュメントから、指定した構造だけを検索し、コピー&ペーストで流用できる

XPathで記述した場合の例
 select="SECTION[TITLE='CDの聞きかた']//解説"

GUI操作



図 4 構造検索例

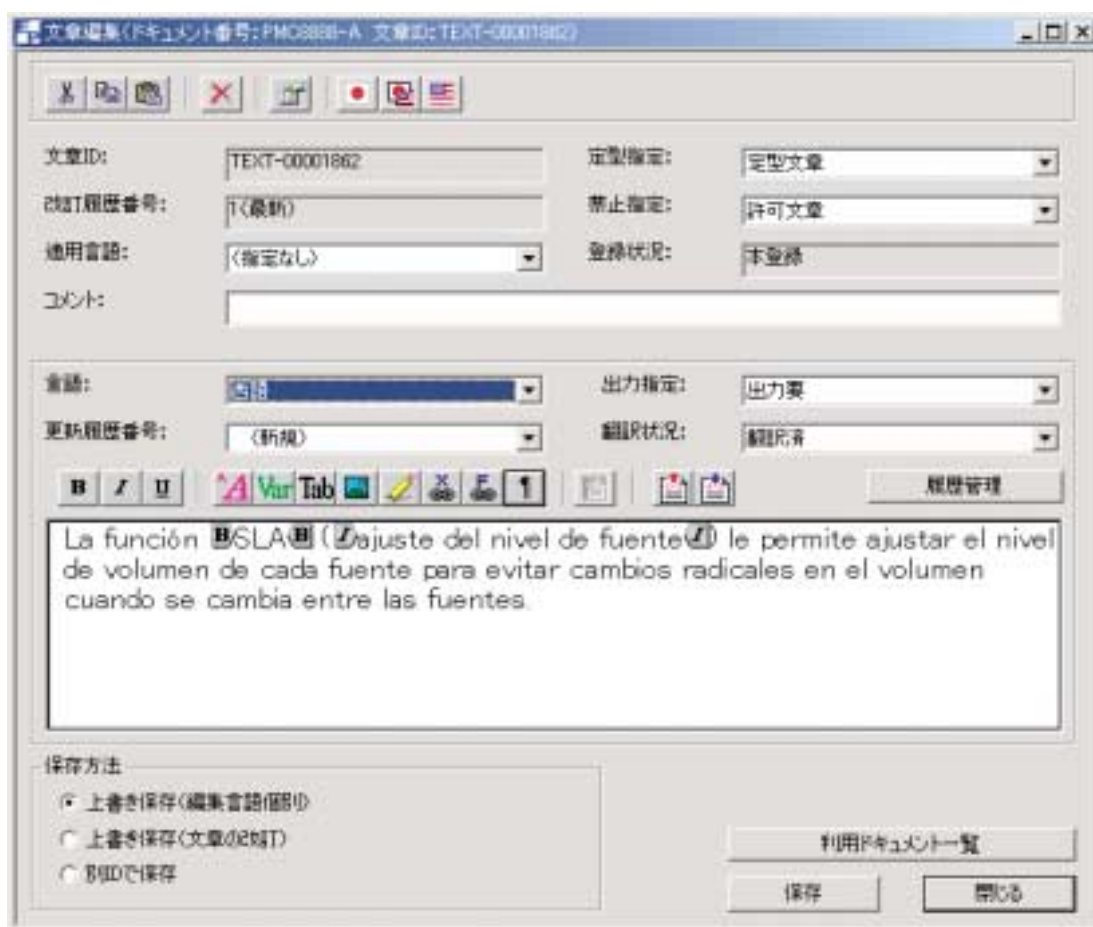


図 5 文章編集画面例

して同じ画面となる。この画面で文章データに関する編集を行う。運用上は、編集するのは主に日本語・英語で、その他の言語は後述する翻訳フローで管理している。

構造を作りながら文章編集を行うため、制作担当者はデータベースに登録している意識はなく、作っているドキュメントが自動的に文章単位でデータベースに登録される。登録された文章は検索が可能になり、他のドキュメントから流用・共有できる。前述の構造検索以外にもテキストベースの全文検索に対応している。

文章は複数のドキュメントから共有されている状態であり、どのドキュメントからでも修正される可能性がある。データベースでは履歴を管理しているため、他の制作担当者の編集により文章の履歴番号が上がった場合は、ドキュメント内の文章リンクが古くなったことを示すマークが表示され、図6に示すように、更新を促す。更新内容の確認結果、反映が好ましくないと判断した場合には、確認済みフラグを立てることによりそのまま使用する選択もできる。

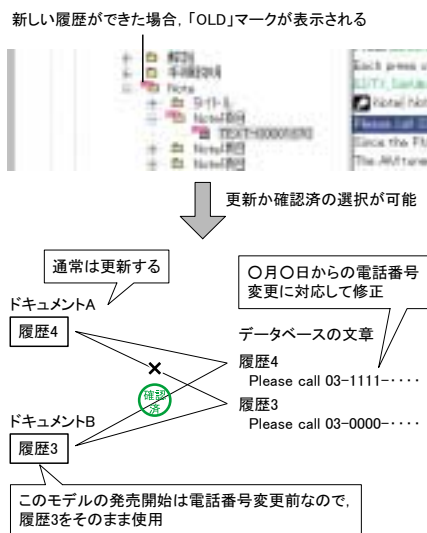


図6 新しい履歴の通知機能

3.3 翻訳管理

取扱説明書では、モデルが違ってても、同じ機能や類似した機能は同じように表現する必要

がある。しかし、モデルによって制作担当者が異なってしまったため、開発日程が重なった場合は、図7に示すように、同じ内容だが違う表現の文章をそれぞれ翻訳依頼すること(二重翻訳)があった。また、たとえ英語が同じ文章でも他言語では異なった翻訳結果になる場合もあり、文章統一の作業には同じかどうかの判断をはじめ多くの工数を費やしてきた。

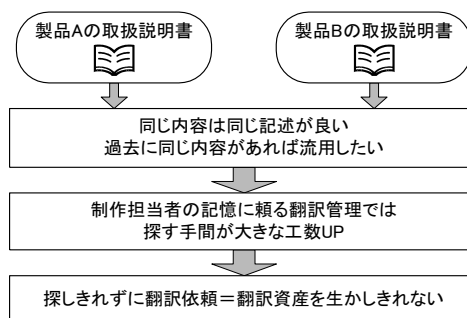


図7 翻訳管理の課題

翻訳は翻訳会社へ依頼するが、翻訳が必要な箇所の選択、抽出、そして翻訳後の反映などは手作業で行っていたため、工数が必要な上に作業ミスが発生する可能性は否定できなかった。翻訳抽出の課題を図8に示す。

図9に示すように、データベースで管理することにより、各言語の翻訳状況もわかり、文章単位に翻訳管理ができる。指定された言語が未翻訳のフラグを持つ文章だけを自動抽出することで、翻訳原稿の完成となる。

なお、翻訳会社ではTRADOSに代表される翻訳支援ツールを使って翻訳作業を効率化しているが、翻訳原稿はTRADOSの入力形式となるファイル形式で出力しており、また翻訳結果もデータベースへ自動取込できるようにしている。

3.4 XML文書生成

構造と文章を組み合わせれば、ドキュメント1冊分のXML文書の完成である。この際に言語を指定すれば、各国語版のドキュメントが生成されることになる。

最終版をXML文書として生成するには、使用

言語すべてが翻訳済みになっている必要がある。このため生成前には、使用言語の翻訳状態や画像ファイルの有無、参照箇所が存在などをまとめてチェックする。図10に表示例を示す。

ただし制作の途中では確認用に未翻訳の文章でも出力することがあるため、未翻訳の文章については代替として英語で出力することも可能としている。

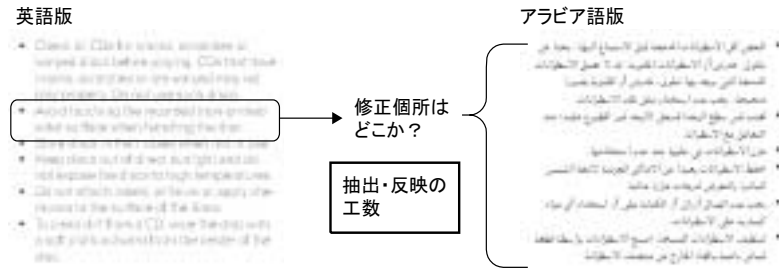


図8 翻訳抽出の課題

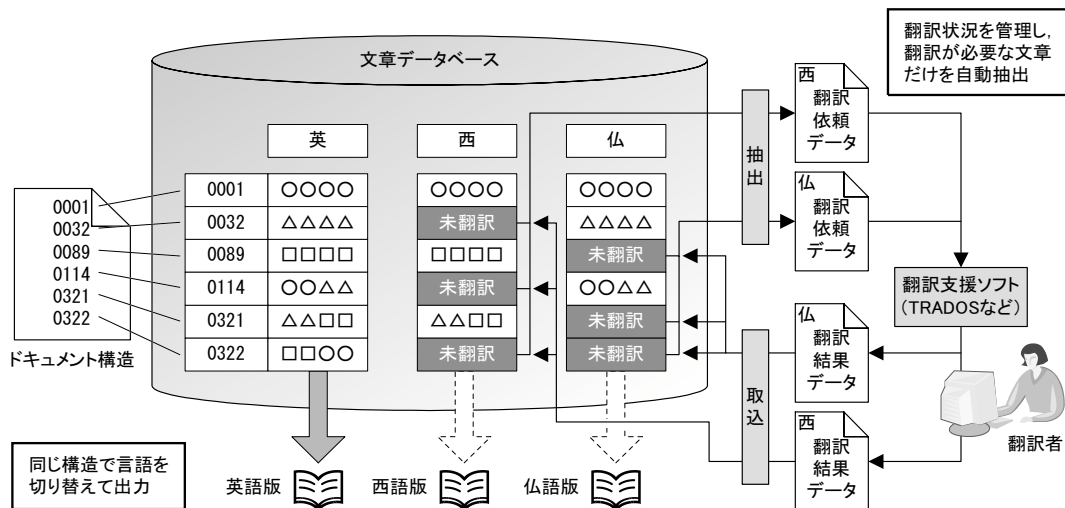


図9 データベースによる翻訳管理

The screenshot shows a software window titled 'スターチェックの結果 (PMC888-A)'. It displays the results of a star check for XML document generation. The table below summarizes the data shown in the screenshot.

#	Asy/パス	整合性	禁止	半登録	更新	参照	画像	翻訳	最新文章
6	/ドキュメント/ドキュメントボディ/オーディオ/ラウドネス/タイトル	○	○	○	○	○	○	×	-
7	/ドキュメント/ドキュメントボディ/オーディオ/ラウドネス/Notes/Note項目	○	○	×	○	○	○	×	-
8	/ドキュメント/ドキュメントボディ/付録/CDケア/タイトル	○	○	○	○	○	○	×	-
9	/ドキュメント/前付け/表紙/音楽	○	○	○	○	○	×	○	-
10	/ドキュメント/前付け/表紙/ロゴ	○	○	○	○	○	×	○	-
11	/ドキュメント/ドキュメントボディ/オーディオ/イコライザー/コピー/手動更新/手動/操作手文/動作ライン/画像/Asy	○	○	×	○	○	○	○	-

Legend: ○: 正常, △: 翻訳中または最新ではないが確認済み, ×: エラー, -: 対象外

図10 XML 文書生成時のチェック結果表示例

生成されたXML 文書は次に述べる自動組版アプリケーションで印刷用のPDF データに変換するが、特に印刷物にする必要がなければ、Web ブラウザで表示することも可能で、実際の表示例を図 11 に示す。

4. 自動組版アプリケーション

XML テクノロジーにおける組版仕様としては XSL ⁵-FO(eXtensible Stylesheet Language - Formatting Objects)が注目されるところだが、ドキュメントレベルの組版機能となるとまだ力不足である。このため、組版エンジンとしてはいくつかの市販パッケージから、次の理由によりアドベント社の 3B2 を選択した。

- ・アプリケーションレベルでUnicode 化されている。
- ・サポート言語が多い(48ヶ国語)。
- ・スクリプトによるレイアウトコントロールが可能である。
- ・状況判断に必要な情報(プロパティ)が多く取得できる。

4.1 単純編集作業の自動化

単純な繰り返し作業である記号系文字のフォント変更、目次や索引の生成、参照ページや参照タイトルの発生、柱や小口見出しの生成は、自動化することでミスを防ぐことができる。ま

た、ボタン名称やディスプレイの表示文字などは変数として扱う仕組みを用意することで、1箇所の修正で済むようにしている。

4.2 レイアウトコントロール

実装した組版スクリプトの例を図 12 に示す。

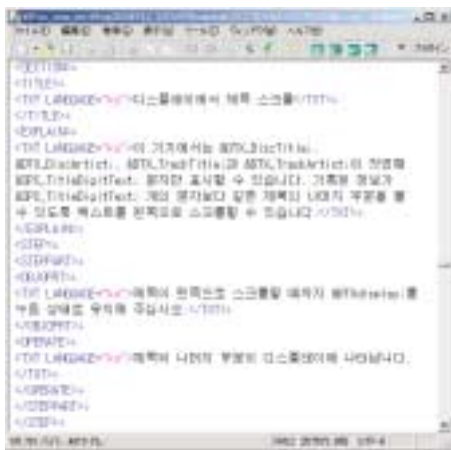
従来から Database Publishing という思想はあったが、構築したのは良いが運用に膨大なデータベース管理工数を必要とする例が多く見受けられた。特に取扱説明書のようなドキュメントの場合、ページ数を抑えるために、自動組版した後に細かなレイアウト調整を手作業で行うことが数多くある。一度レイアウト調整した後は、記載内容の修正も手作業となるため、度重なる仕様変更の修正結果をデータベースにフィードバックするメンテナンス工数が課題として認識されていたのである。

今回のシステムでは、自動組版の過程にスクリプトを組み込むことで状況判断に応じた詰め処理も可能となり、組版の精度を高めることができた。このため手作業によるレイアウト調整を排除し、データベース自体を修正してから再度自動組版するフローで運用が可能である。

5. まとめ

このシステムは現在、海外向けカーオーディオ製品の取扱説明書の制作に利用している。自

生成されたXML形式の取扱説明書(韓国語)



スタイルを定義すればブラウザ表示も可能



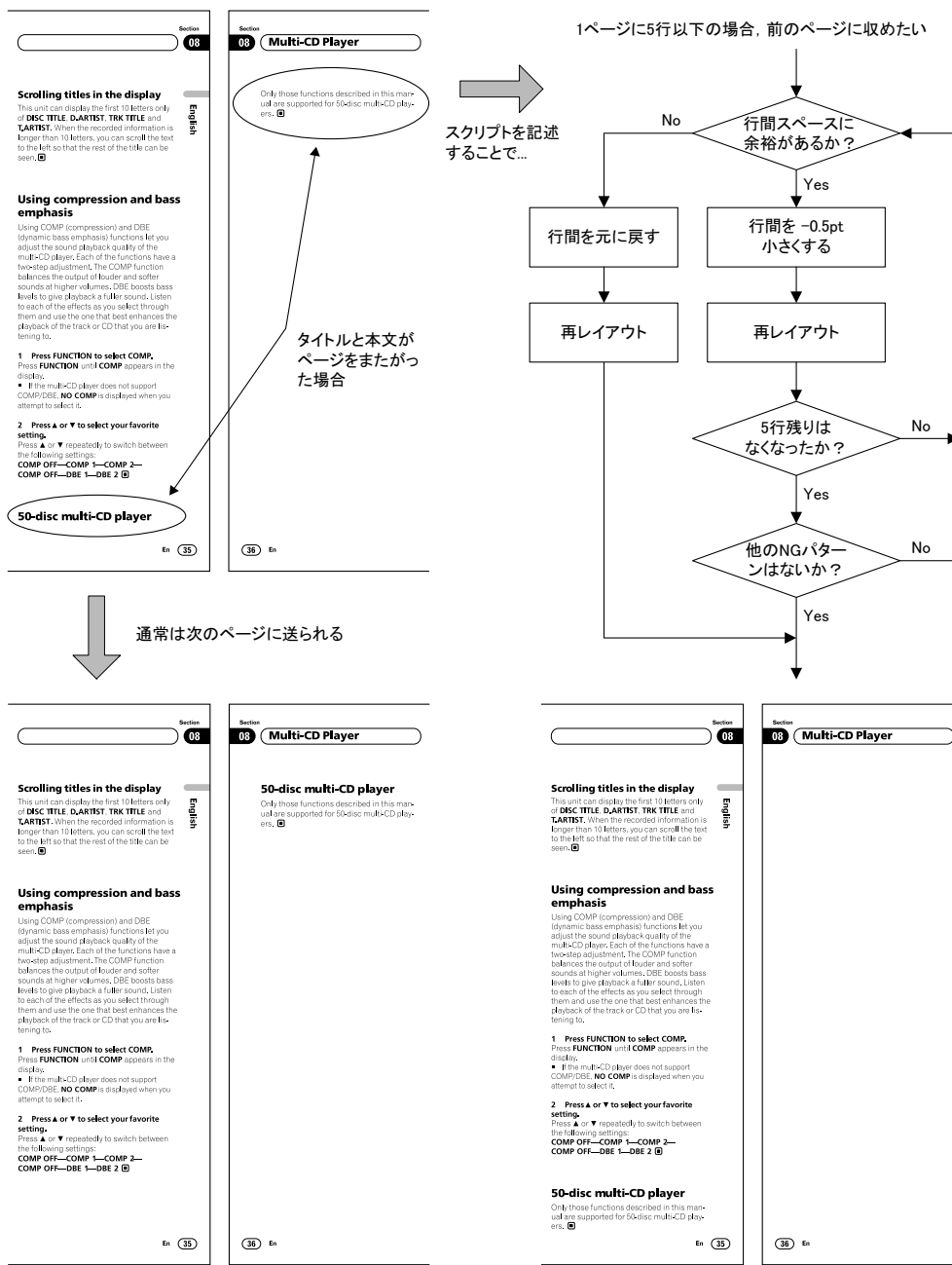
図 11 XML 取扱説明書

動組版の性質上、レイアウトコントロールには限界があり柔軟なレイアウト変更には課題が残るが、記載情報の標準化や編集工数の削減などXMLで管理できるメリットは大きい。

製品が多機能・高機能になるほど、製品とユーザをつなぐコミュニケーションツールとし

て取扱説明書の持つ役割は重要になってくる。検索性やインタラクティブ性といったわかりやすさに直結する観点から電子メディアへも展開されており、紙にこだわらない取扱説明書も現在では珍しくない。

今後は、わかりやすさ品質の向上により多く



前のページに収めたい場合、手作業となる
↓
記載内容に修正が入った場合、再レイアウト調整を避けるため、組版アプリケーションで直接修正してしまう
↓
修正内容をデータベースにフィードバックする必要がある

前のページに収める作業が自動で行われる
↓
記載内容に修正が入った場合、データベースを修正してから再度組版しても、手作業でレイアウト調整をする必要はない
↓
データベースを常に最新に保つことができる

図 12 組版スクリプト概要

の時間をかけることが可能な制作フローを構築するとともに、XML という再利用しやすいデータ形式の利点を生かして、公開情報のワンソース・マルチユース化を進めていく予定である。

6. 謝辞

本システムの開発にあたり、日立ソフトウェアエンジニアリング(株)ならびにアルテック(株)の関係各位に多大なるご協力をいただきました。この紙面をお借りして厚くお礼申し上げます。

用語解説

1. DTP(DeskTop Publishing):

パーソナルコンピュータを使って出版物を作成すること。文字や図版の入力から始まって、ページ全体のレイアウトから印刷の前段階(Prepress)まで含まれることもある。そのため、DeskTop Prepress の略で使われることもある。

2. XML(eXtensible Markup Language):

Internet 上での SGML の利用を容易にすることを目的として設計されたマークアップ言語。HTML のような固定のマークアップ方法だけではなく、文書独自のマークアップ方法を定義できるようになる。この XML の標準化は、W3C で行なわれている。

3. DTD(Document Type Definition):

SGML で使われる文書型定義。SGML は汎用マークアップ言語であり、どのような要素をどのようなタグを用いてマークアップするかは、用途に応じて任意に決められる。そして、それを決めるのが DTD の役目でもある。

注: XML でも同様の目的で利用することができる。

4. XPath(XML Path Language):

XML 文書構造にアドレッシングするための言語。

5. XSL(eXtensible Stylesheet

Language):

XML 文書のデータ内容(コンテンツ)を画面や紙の上にフォーマットするためのスタイルシート言語。

注: XSL の開発の初期段階では、XML 文書を他のデータ形式に変換する部分(Transformations)とフォーマットする部分(Formatting Objects)が両方とも含まれていたが、現在では変換言語は XSLT という独立した規格となっており、XSL と言えばフォーマット言語(XSL-F0)を指すようになっている。

参考文献

- (1) アスキー デジタル用語辞典:
<http://yougo.ascii24.com/>
- (2) Elliotte Rusty Harold 著, 藤本叔子 訳:
「XML バイブル」
- (3) Neil Bradley 著, 安藤 慶一 訳:
「XSLT 完全活用マニュアル」

筆者

吉田 樹治(よしだ しげじ)

所属: (株) パイオニアメディアクリエイティブ
開発部

入社年月: 1990 年 4 月

主な経歴: マニュアル制作関連のシステム開発

間島 誠也(まじま せいや)

所属: (株) パイオニアメディアクリエイティブ
開発部

入社年月: 1994 年 4 月

主な経歴: マニュアル制作関連のシステム開発

大橋 則行(おおはし のりゆき)

所属: (株) パイオニアメディアクリエイティブ
川越 マニュアルセンター

入社年月: 1985 年 4 月

主な経歴: マニュアル制作関連の管理業務